

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-020501

(43)Date of publication of application : 21.01.2000

(51)Int.Cl.

G06F 17/10

G06F 15/16

(21)Application number : 10-188840

(71)Applicant : TOSHIBA CORP

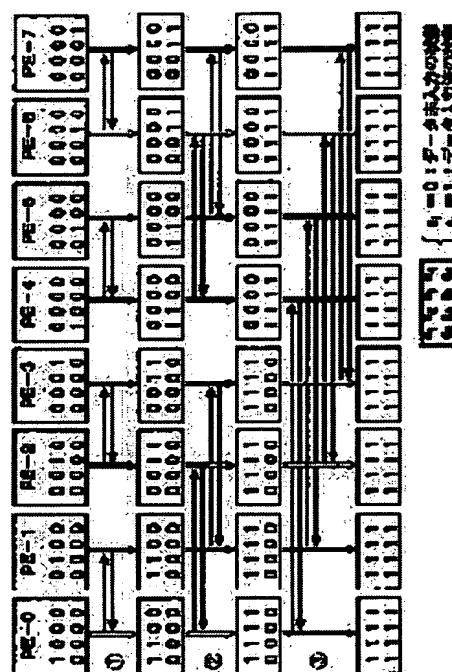
(22)Date of filing : 03.07.1998

(72)Inventor : UEMATSU MIKIO

(54) PARALLEL COMPUTER SYSTEM AND COMMUNICATION METHOD BETWEEN ARITHMETIC PROCESSING UNITS**(57)Abstract:**

PROBLEM TO BE SOLVED: To efficiently collect data distributively processed by respective arithmetic processing units(APU) in a parallel computer.

SOLUTION: In the case of collecting data arrays divided into 2^n small arrays and distributed/computed by respective APUs as one array in the parallel computer system provided with 2^n APUs to which identification(ID) numbers 0, 1,..., $2^n - 1$ are given, individual storage means and communication means, a number N' obtained by inverting the value of 2^i -th digit of an ID number N expressed by binary notation is allowed to correspond to the ID number N and operation (i) for mutually transmitting/receiving the arithmetic processing results of data arrays between the APU of the ID number N and the APU of the ID number N' is successively executed from $i=0$ to $i=n-1$. When $j>0$, arithmetic processing results obtained up to operation (j-1) are transmitted/received between the APUs of the ID numbers N , N' in addition to the arithmetic processing results obtained by respective APUs at the time of operation (i).

**LEGAL STATUS**

[Date of request for examination]

01.06.2004

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-20501

(P 2 0 0 0 - 2 0 5 0 1 A)

(43) 公開日 平成12年1月21日(2000.1.21)

(51) Int. Cl. ⁷	識別記号	F I	テマコード (参考)
G06F 17/10		G06F 15/31	Z 5B045
15/16	390	15/16	390 Z 5B056

審査請求 未請求 請求項の数 7 O L (全17頁)

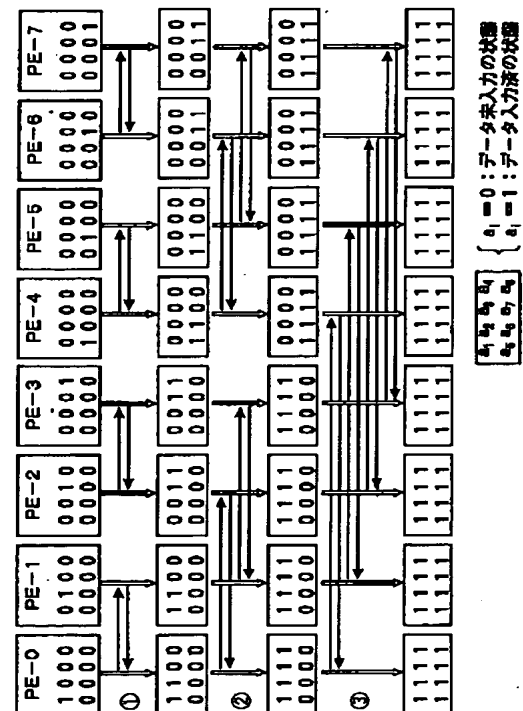
(21) 出願番号	特願平10-188840	(71) 出願人	000003078 株式会社東芝 神奈川県横浜市幸区堀川町72番地
(22) 出願日	平成10年7月3日(1998.7.3)	(72) 発明者	上松 幹夫 神奈川県横浜市磯子区新杉田町8番地 株 式会社東芝横浜事業所内
		(74) 代理人	100083161 弁理士 外川 英明
		F ターム(参考)	5B045 AA07 BB02 BB28 BB47 GG02 GG12 5B056 AA04 BB45 DD14 FF05

(54) 【発明の名称】 並列計算機システム及びその演算処理装置間の通信方法

(57) 【要約】

【課題】 並列計算機の各演算処理装置で分散処理されたデータを効率的に集結する。

【解決手段】 識別番号 $0, 1, \dots, 2^n - 1$ が付与された 2^n 台の演算処理装置と個別記憶装置及び通信手段を備えた並列計算機システムで、 2^n 個の小配列に分割して各演算処理装置に分配／演算処理されたデータ配列を1つの配列に集結する際に、識別番号 N に対し2進法で表した識別番号 N の 2^i の位の数を反転させた番号 N' を対応させ、識別番号 N の演算処理装置と識別番号 N' の演算処理装置の間でデータ配列の演算処理結果を相互に送受信する操作 i を $i = 0$ から $i = n - 1$ まで順次行う。この際、 $j > 0$ なる j に対しては、操作 i の際に、識別番号 N, N' の演算処理装置間で各演算処理装置による演算処理結果に加えて操作 $(j - 1)$ までで得られた演算処理結果を送受信する。



【特許請求の範囲】

【請求項1】 固有の識別子を有する少なくとも 2^n 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 2^n 個の小配列に分割して 2^n 台の演算処理装置に分配され各演算処理装置で演算処理されたデータ配列を再び1つの配列に集結する際に、 2^n 台の演算処理装置に識別番号0, 1, ..., $2^n - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi = 0からi = n - 1まで順次行い、j > 0なるjに対しては、操作jの際に、識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j - 1)までで得られた演算処理結果を送受信することにより 2^n 台の演算処理装置間でn回の操作でデータ配列を集結させることを特徴とする並列計算機システム。

【請求項2】 固有の識別子を有する $(2^n + k)$ 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 $(2^n + k)$ 個の小配列に分割して $(2^n + k)$ 台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、前記 $(2^n + k)$ 台の演算処理装置に個別記憶手段及び通信手段を備えた $(2^n - k)$ 台の演算処理装置を加えた 2^{n+1} 台からなる演算処理装置群を形成し、この演算処理装置群を構成する 2^{n+1} 台の演算処理装置に識別番号0, 1, ..., $2^{n+1} - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi = 0からi = mまで順次行い、j > 0なるjに対しては、操作jの際に、 $N \leq 2^n + k$ なる識別番号Nの演算処理装置からはその演算処理装置の演算処理結果及び操作(j - 1)までで得られた演算処理結果を送信し、 $N > 2^n + k$ なる識別番号Nの演算処理装置からは操作(j - 1)までで得られた演算処理結果を送信することにより $(2^n + k)$ 台の演算処理装置において(m + 1)回の操作でデータ配列を集結させることを特徴とする並列計算機システム。

【請求項3】 固有の識別子を有する $(2^n + k)$ 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機

システムにおいて、 $(2^n + k)$ 個の小配列に分割して $(2^n + k)$ 台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、この $(2^n + k)$ 個のデータ配列に $(2^n - k)$ 個の空の小配列を追加することで前記データ配列を小配列 2^{n+1} 個分の配列に拡張し、前記 $(2^n + k)$ 台の演算処理装置に、個別記憶手段及び通信手段を備えた $(2^n - k)$ 台の演算処理装置を加えた 2^{n+1} 台からなる演算処理装置群を形成し、この演算処理装置群を構成する 2^{n+1} 台の演算処理装置に識別番号0, 1, ..., $2^{n+1} - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi = 0からi = mまで順次行い、j > 0なるjに対して、操作jの際に、識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j - 1)までで得られた演算処理結果を送受信することにより $(2^n + k)$ 台の演算処理装置において(m + 1)回の操作でデータ配列を集結させることを特徴とする並列計算機システム。

【請求項4】 $n > m$ なるn, mについて、固有の識別子を有する $(2^n + 2^m)$ 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 $(2^n + 2^m)$ 個の小配列に分割して $(2^n + 2^m)$ 台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、前記 $(2^n + 2^m)$ 台の演算処理装置を 2^n 台からなるグループG₁と 2^m 台からなるグループG₂に分割し、また前記データ配列を初めの 2^n 個の小配列からなる配列A₁とその後の 2^m 個の小配列からなる配列A₂の2つに分割し、この配列A₁, A₂をそれぞれグループG₁, G₂と対応づけて分配、演算処理を行い、グループG₁の 2^n 台の演算処理装置に識別番号0, 1, ..., $2^n - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi = 0からi = n - 1まで順次行い、j > 0なるjに対して、操作jの際に識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j - 1)までで得られた演算処理結果を送受信することによりグループG₁内で配列A₁を集結させる第1の工程と、グループG₂の 2^m 台の演算処理装置に識別番号0, 1, ..., $2^m - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識

別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi=0からi=n-1まで順次行い、j>0なるjに対して、操作jの際に識別番号N、N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信することによりグループG_i内で配列A_iを集結させる第2の工程と、グループG_iからグループG_jの各演算処理装置に配列A_iを、グループG_jからグループG_iの各演算処理装置に配列A_jを送信する第3の工程とを有し、第1の工程と第2の工程を並列に実行した後に第3の工程を行なうことにより(2ⁿ+2ⁿ)台の演算処理装置においてデータ配列を集結させることを特徴とする並列計算機システム。

【請求項5】 固有の識別子を有する複数の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備えた並列計算機システムにおいて、

【数1】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

個の小配列(但し、n₁>n₂>n₃>...>n_k≥0)に分割して

【数2】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、これらの演算処理装置のうち

【数3】

$$2^{n_1}, 2^{n_2}, 2^{n_3}, \dots, 2^{n_k}$$

台をそれぞれグループG₁, G₂, ..., G_kとしてk個のグループに分割するとともに、前記小配列のうち

【数4】

$$2^{n_1}, 2^{n_2}, 2^{n_3}, \dots, 2^{n_k}$$

個の小配列をそれぞれ配列A₁, A₂, ..., A_kとしてk個の配列に分割し、このk個の配列とk個のグループG₁, G₂, ..., G_kとを1対1に対応づけて分配・演算処理を行い、1≤p≤kなる各pに対し、グループG_pの(2のn_p乗)台の演算処理装置に識別番号0,

1, ...を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの2ⁱの位の数値を反転させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi=0からi=n-1まで順次行い、j>0なるjに対して、操作jの際に識別番号N、N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信する

ことによりグループG_p内の演算処理装置でデータ配列A_pを集結させるグループ内工程pを実行し、グループ内工程(k-1)が終了した後、グループG_kの演算処理装置から配列A_kの演算結果をグループG_{k-1}の演算処理装置に送信するグループ間工程kを実行し、次に、グループG_kの各演算処理装置に集結された配列A_kの演算結果を、グループG_kの演算処理装置からq>pなる全てのqに対しグループG_qに属する各演算処理装置に送信するとともに、グループG_qの演算処理装置から、グループG_q自身の演算結果である配列A_q及びグループG_{q+1}の演算処理装置から受信した配列A_{q-1}, ..., A_kの演算結果をグループG_{q-1}の演算処理装置に送信するグループ間工程pを、p=k-1からp=2までpに関して降順に実行することにより、

【数5】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

台の演算処理装置においてデータ配列を集結させることを特徴とする並列計算機システム。

【請求項6】 請求項4または5記載の並列計算機システムを用いて演算処理装置のグループ間でのデータ交換を行う場合、p>qなるp, qについて、2^p台の演算処理装置からなるグループG_pで集結され共有されているデータ配列Aと、2^q台の演算処理装置からなるグループG_qで集結され共有されているデータ配列Bとを、グループG_p, G_q間で相互に送受信する際に、グループG_pのなかから選択される2^q台の演算処理装置をグループG_qの各演算処理装置と1対1に対応させてグループG_qの各演算処理装置にデータ配列Aを送信する操作を並列に実施するとともに、グループG_pを、それぞれが2^{p-q}台の演算処理装置からなる小グループα₁, α₂, ..., α_r (r=2^q)に分割して、各々の小グループとグループG_qのr台の各演算処理装置とを1対1に対応させ、小グループα_iのなかから選択される1台の演算処理装置に対して、小グループα_iに対応するグループG_qの演算処理装置からデータ配列Bを送信した後、小グループα_iの演算処理装置間でデータ配列Bを送受信する操作iを、1≤i≤rなるiに関して並列に実行することにより、2^p台の演算処理装置と2^q台の演算処理装置にデータ配列Aとデータ配列Bを共有させることを特徴とする並列計算機システム。

【請求項7】 請求項1乃至6のいずれか記載の並列計算機システムを用いて2台の演算処理装置間でデータを交換する工程は、演算処理装置の識別番号の大きい方から小さい方にデータを送る第1の送信工程と、演算処理装置の識別番号の小さい方から大きい方にデータを送る第2の送信工程とからなり、この第1の送信工程と第2の送信工程のうちから選択される1工程を先に行った後、続いて他の1工程を行うことを特徴とする並列演算機システムの演算処理装置間の通信方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、通信手段および個別記憶装置を備えた多数の演算処理装置からなり、特に並列計算を目的とした並列計算機システム及びその演算処理装置の通信方法に関する。

【0002】

【従来の技術】原子力施設をはじめとする大規模な施設の設計においては、例えば遮蔽設計などにおける放射線挙動計算、炉心設計における炉心性能予測解析などの大規模な計算がかなりの頻度で要求される。この要求に応えるためには大幅な計算速度の向上が必要である。このため最近では、通信手段と個別の記憶装置を備えた多数の演算処理装置を用いて、1台の演算処理装置しか持たない計算機を使用していたのでは得られないような高速度で、解析を行うことが考案されている。

【0003】例えば炉心設計であれば、原子炉の炉心を複数の燃料集合体からなる幾つかのセグメントに分割し、それぞれのセグメントを1つの演算処理装置に対応させて、出力計算と熱水力計算を各々の演算処理装置で並列に計算させる。セグメント間での中性子束の流出入およびチャンネル間の冷却材の圧力バランスを解析する際には、前記通信手段によりセグメント境界の中性子

束、各チャンネルの圧力損失のデータを演算処理装置間でやり取りすることで、空間的に連続した解析が行われる。

【0004】また、遮蔽設計であれば、例えば原子炉の炉心、冷却材、遮蔽体などを含む全体系を幾つかの小領域に分割し、それぞれの小領域を1つの演算処理装置に対応させて、放射線束分布計算を各々の演算処理装置で並列に計算させる。小領域間での中性子束の流出入を解析する際には、前記通信手段により小領域境界の中性子束のデータを演算処理装置間でやり取りすることで、空間的に連続した解析が行われる。

【0005】

【発明が解決しようとする課題】複数の演算処理装置を用いて並列に計算を行わせる際に、演算処理装置間の通信を行うことなく全く独立に計算を進めることができる例はまれであり、通常は演算処理装置間の通信を行いながら計算を進める。たとえば、4行4列の行列A、Bの掛け算を4台の演算処理装置で実施して4行4列の行列Cを求める場合を考える。A、B、Cの要素をそれぞれ a_{ij} 、 b_{ij} 、 c_{ij} で以下のように表記する。

【0006】

【数6】

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \\ b_{31} & b_{32} & b_{33} & b_{34} \\ b_{41} & b_{42} & b_{43} & b_{44} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}$$

このとき、4台の演算処理装置のうちの1台においては例えば、

【0007】

【数7】

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

のように計算が行われる。

【0008】この例から明らかなように、演算に使う側 (a_{ij} または b_{ij}) については行或いは列全体についての要素のデータが必要である。また、演算の結果として得られる c_{ij} の方は、各々の演算処理装置に於いては部分的にしかデータが得られない。このことは、例えば次のステップで行列Cと行列Aの掛け算を行う必要が生じたとき、計算で得られた要素だけではデータに不足が生じることを意味する。したがって、 $A \times B = C$ の計算を実施した後で残りの部分、上の式で言えば行列Cの少なくとも第1行と第2行のデータ及び第1列と第2列のデータは満たされた状態にしておかねばならない。

【0009】これらの問題を一般化すると次のようになる。 $(n \times k)$ 個からなる配列 $X(nk)$ があり、これが n 台の演算処理装置に分割され、例えば識別番号1の演算処理装置では $X(1)$ 、 $X(2)$ 、 \dots 、 $X(k)$ 、識別番号2の演算処理装置では $X(k+1)$ 、 $X(k+2)$ 、 \dots 、 $X(2k)$ の計算結果を持っているものとする。この状態から n 台の演算処理装置の間で通信を行うことにより、 n 台の演算処理装置が配列 $X(nk)$ の計算結果を持っている状況を作る操作が必要となることがある。

【0010】このときの通信は1対1であることが通信手段上の条件である。すなわち、例えば演算処理装置1から演算処理装置2にデータを転送する際には、演算処理装置2は演算処理装置1からデータを受けとる態勢になければならないのであって、このとき演算処理装置2が他の処理、例えば演算処理装置3にデータを転送しようとしたり演算処理装置4からデータを受けようとしたりすると、通信は失敗して計算は中断することとなる。通信が滞りなく行われるには送信側と受信側の混乱がないように通信の順序を予め決めておく必要がある。

【0011】4台の演算処理装置を使う場合を例にとれば、容易に考えられる方法として次のものが挙げられ

10

20

30

40

50

る。以下、表記を簡略化するため演算処理装置 1, 2, 3, 4 をそれぞれ # 1, # 2, # 3, # 4 と書く。

(1) 送信-受信を 1 つずつ順次行う方法

- [1] # 1 の計算結果 → # 2, [2] # 1 の計算結果 → # 3,
- [3] # 1 の計算結果 → # 4, [4] # 2 の計算結果 → # 1,
- [5] # 2 の計算結果 → # 3, [6] # 2 の計算結果 → # 4,
- [7] # 3 の計算結果 → # 1, [8] # 3 の計算結果 → # 2,
- [9] # 3 の計算結果 → # 4, [10] # 4 の計算結果 → # 1,
- [11] # 4 の計算結果 → # 2, [12] # 4 の計算結果 → # 3

を順次実行する。

【0012】ここで、[1], [2], [3], ... は処理のステップの番号を示す。演算処理装置を N 台、1 台に割り

当てられたデータ量を w とすれば、通信回数は

$$2 \times N \times C_2 = N(N-1)$$

であり、データ移動量は

$$2w \times N \times C_2 = wN(N-1)$$

・データ集結

- [1] # 2 の計算結果 → # 1, [2] # 3 の計算結果 → # 1,
- [3] # 4 の計算結果 → # 1, ... を順次実行。

1 に全データが揃う。

・全データ配布

- [1] # 1 → # 2, [2] # 1 → # 3, [3] # 1 → # 4, ...

を順次実行。配列全体を # 2, # 3, # 4 に送信する。

【0014】この場合の通信回数は $2(N-1)$ 回、データ移動量は集結時に $(N-1)w$ 、配布時に $N(N-1)w$ である。この方法は (1) の方法に比べて通信回数は少ないが、全データ配布時に送信されるデータ

- [1] # 1 の計算結果 → # 2, # 3 の計算結果 → # 4 を同時に実行。
- [2] # 1 の計算結果 → # 3, # 2 の計算結果 → # 4 を同時に実行。
- [3] # 1 の計算結果 → # 4, # 2 の計算結果 → # 3 を同時に実行。
- [4] # 2 の計算結果 → # 1, # 4 の計算結果 → # 3 を同時に実行。
- [5] # 3 の計算結果 → # 1, # 4 の計算結果 → # 2 を同時に実行。
- [6] # 4 の計算結果 → # 1, # 3 の計算結果 → # 2 を同時に実行。

【0016】この通信方法によれば、通信が重複することも衝突することもなく、全データが 4 台の演算処理装置に行き渡る。演算処理装置が N 台であれば通信回数は $2(N-1)$ 、データ移動量は $2(N-1)w$ である。N=4 であれば通信に要する時間は前述の (1) の方法の半分である。N が大きくなるとともに差は広がる。

【0017】(4) 演算処理装置の Binary tree により代表の演算処理装置にデータを集めた後、各演算処理装置に配布する。これは (2) の方法を改良したもので、例えば次のように行う。

・データ集結

- [1] # 2 の計算結果 → # 1, # 4 の計算結果 → # 3 を同時に実行。

- [2] # 3 に集結された計算結果 → # 1

・全データ配布

- [3] # 1 → # 3

- [4] # 1 → # 2, # 3 → # 4 を同時に実行。

である。N=4 ならば通信回数は上述の 12 回である。この方法によれば、時間はかかるが通信上の混乱は避けられる。なお、 C_2 は p 個の要素から q 個の要素を選ぶ組合せの数を示す。

【0013】(2) 代表の演算処理装置にデータを集めた後、各演算処理装置に配布する。

タ量が多い点が短所である。

【0015】また、通信の効率化を図った手法として次のものがある。

(3) 演算処理装置の 1 対 1 の組み合わせに対して並列・網羅的に通信を行う。これは (1) の方法を改良したもので、例えば次のように行う。

【0018】この方法によれば、演算処理装置が N 台であれば、通信回数は $2 \times \log_2 N$ 回、データ通信量は、集結時に $(N-1)w$ 、配布時に $Nw \log_2 N$ である。N=4 であれば通信回数は (2) の方法の $2/3$ 、データ移動量は (2) の方法の $11/15$ である。N が大きくなるとともに差は広がる。

【0019】(3) の方法は (4) の方法に比べてデータ移動量は少ないが通信回数が多いため、扱う配列が小さい場合には適していない。(4) の方法は通信回数は少ないが、データ移動量が多いため、巨大な配列を扱う場合には適していない。

【0020】よって、データ移動量と通信回数がともに最適化された、あらゆる条件に対して適用可能な一般化された手法が必要である。本発明は、このような点を考慮してなされたもので、通信によるデータの授受を並列に行えるようにすることで、演算処理装置間の通信回数およびデータの授受の際の待ち時間を最小限に抑えて高

速化を図ることができる並列計算機システム及びその演算処理装置間の通信方法を提供することを目的とする。

【0021】

【課題を解決するための手段】上記目的を達成するため、本発明の請求項1記載の発明は、固有の識別子を有する少なくとも 2^n 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 2^n 個の小配列に分割して 2^n 台の演算処理装置に分配され各演算処理装置で演算処理されたデータ配列を再び1つの配列に集結する際に、 2^n 台の演算処理装置に識別番号0, 1, ..., $2^n - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を反転させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iを $i=0$ から $i=n-1$ まで順次行い、 $j>0$ なるjに対しては、操作jの際に、識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信することにより 2^n 台の演算処理装置間でn回の操作でデータ配列を集結させることを特徴とする。

【0022】また、請求項2記載の発明は、固有の識別子を有する $(2^n + k)$ 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 $(2^n + k)$ 個の小配列に分割して $(2^n + k)$ 台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、前記 $(2^n + k)$ 台の演算処理装置に個別記憶手段及び通信手段を備えた $(2^n - k)$ 台の演算処理装置を加えた 2^{n+1} 台からなる演算処理装置群を形成し、この演算処理装置群を構成する 2^{n+1} 台の演算処理装置に識別番号0, 1, ..., $2^{n+1} - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を反転させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iを $i=0$ から $i=m$ まで順次行い、 $j>0$ なるjに対しては、操作jの際に、 $N \leq 2^n + k$ なる識別番号Nの演算処理装置からはその演算処理装置の演算処理結果及び操作(j-1)までで得られた演算処理結果を送信し、 $N > 2^n + k$ なる識別番号Nの演算処理装置からは操作(j-1)までで得られた演算処理結果を送信することにより $(2^n + k)$ 台の演算処理装置において(m+1)回の操作でデータ配列を集結させることを特徴とする。

【0023】また、請求項3記載の発明は、固有の識別子を有する $(2^n + k)$ 台の演算処理装置と、これら各

演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 $(2^n + k)$ 個の小配列に分割して $(2^n + k)$ 台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、この $(2^n + k)$ 個のデータ配列に $(2^n - k)$ 個の空の小配列を追加することで前記データ配列を小配列 2^{n+1} 個分の配列に拡張し、前記 $(2^n + k)$ 台の演算処理装置に個別記憶手段及び通信手段を備えた $(2^n - k)$ 台の演算処理装置を加えた 2^{n+1} 台からなる演算処理装置群を形成し、この演算処理装置群を構成する 2^{n+1} 台の演算処理装置に識別番号0, 1, ..., $2^{n+1} - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を反転させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iを $i=0$ から $i=m$ まで順次行い、 $j>0$ なるjに対して、操作jの際に、識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信することにより $(2^n + k)$ 台の演算処理装置において(m+1)回の操作でデータ配列を集結させることを特徴とする。

【0024】また、請求項4記載の発明は、 $n>m$ なるn, mについて、固有の識別子を有する $(2^n + 2^m)$ 台の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備え、この通信手段により各演算処理装置間でデータの授受を行う並列計算機システムにおいて、 $(2^n + 2^m)$ 個の小配列に分割して $(2^n + 2^m)$ 台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、前記 $(2^n + 2^m)$ 台の演算処理装置を 2^n 台からなるグループG₁と 2^m 台からなるグループG₂に分割し、また前記データ配列を初めの 2^n 個の小配列からなる配列A₁とその後の 2^m 個の小配列からなる配列A₂の2つに分割し、この配列A₁, A₂をそれぞれグループG₁, G₂と対応づけて分配、演算処理を行い、グループG₁の 2^n 台の演算処理装置に識別番号0, 1, ..., $2^n - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を反転させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iを $i=0$ から $i=n-1$ まで順次行い、 $j>0$ なるjに対して、操作jの際に識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信することによりグループG₁内でデータ配列を集結させる第1の工程と、グループG₂の 2^m 台の演算処理装置に識

11

別番号0, 1, ..., $2^n - 1$ を付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操作iをi=0からi=n-1まで順次行い、j>0なるjに対して、操作jの際に識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信することによりグループG_i内でデータ配列を集結させる第2の工程と、グループG_iからグループG_jの各演算処理装置に配列A_iを、グループG_jからグループG_iの各演算処理装置に配列A_jを送信する第3の工程とを有し、第1の工程と第2の工程を並列に実行した後に第3の工程を行なうことにより($2^n + 2^n$)台の演算処理装置においてデータ配列を集結させることを特徴とする。

【0025】また、請求項5記載の発明は、固有の識別子を有する複数の演算処理装置と、これら各演算処理装置に各々対応する個別記憶装置および通信手段とを備えた並列計算機システムにおいて、

【数8】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

個の小配列(但し、 $n_1 > n_2 > n_3 > \dots > n_k \geq 0$)に分割して

【数9】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

台の演算処理装置に分配・演算処理されたデータ配列を再び1つの配列に集結する際に、これらの演算処理装置のうち

【数10】

$$2^{n_1}, 2^{n_2}, 2^{n_3}, \dots, 2^{n_k}$$

台をそれぞれグループG₁, G₂, ..., G_kとしてk個のグループに分割するとともに、前記小配列のうち

【0026】

【数11】

$$2^{n_1}, 2^{n_2}, 2^{n_3}, \dots, 2^{n_k}$$

個の小配列をそれぞれ配列A₁, A₂, ..., A_kとしてk個の配列に分割し、このk個の配列とk個のグループG₁, G₂, ..., G_kとを1対1に対応づけて分配、演算処理を行い、 $1 \leq p \leq k$ なる各pに対し、グループG_pの(2^n の n_p 乗)台の演算処理装置に識別番号0, 1, ..., 付与し、識別番号Nの演算処理装置に対し2進法で表した識別番号Nの 2^i の位の数を取反させた番号N'を識別番号とする演算処理装置を対応させ、前記データ配列の演算処理結果を識別番号Nの演算処理装置と識別番号N'の演算処理装置の間で相互に送受信する操

12

作iをi=0からi=n-1まで順次行い、j>0なるjに対して、操作jの際に識別番号N, N'の演算処理装置間で各演算処理装置による演算処理結果に加えて操作(j-1)までで得られた演算処理結果を送受信することによりグループG_i内の演算処理装置でデータ配列A_iを集結させるグループ内工程pを実行し、グループ内工程(k-1)が終了した後、グループG_kの演算処理装置から配列A_kの演算結果をグループG_{k-1}の演算処理装置に送信するグループ間工程kを実行し、次に、グループG_kの各演算処理装置に集結された配列A_kの演算結果を、グループG_kの演算処理装置からq>pなる全てのqに対しグループG_kに属する各演算処理装置に送信するとともに、グループG_kの演算処理装置から、グループG_k自身の演算結果である配列A_k及びグループG_{k+1}の演算処理装置から受信した配列A_{k+1}, ..., A_kの演算結果をグループG_{k+1}の演算処理装置に送信するグループ間工程pを、p=k-1からp=2までpに関して降順に実行することにより、

【0027】

【数12】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

台の演算処理装置においてデータ配列を集結させることを特徴とする。

【0028】なお、この際には、k個のグループ内工程1, 2, ..., kを並列に実行し、 $1 \leq s \leq k-1$ なるsに対して、グループ内工程sが終了した時点で順次グループ間工程(s+1)を実行することで、全体の通信に要する時間をさらに短縮することができる。

30

【0029】また、請求項6記載の発明は、請求項4または5記載の並列計算機システムを用いて演算処理装置のグループ間でのデータ交換を行う場合、 2^n 台の演算処理装置からなるグループG_kで集結され共有されているデータ配列Aと、 2^n 台の演算処理装置($p > q$)からなるグループG_qで集結され共有されているデータ配列Bとを、グループG_k, G_q間で相互に送受信する際に、グループG_kのなかから選択される 2^n 台の演算処理装置をグループG_qの各演算処理装置と1対1に対応させてグループG_qの各演算処理装置にデータ配列Aを送信する操作を並列に実施するとともに、グループG_kを、それぞれが 2^{p-q} 台の演算処理装置からなる小グループ $\alpha_1, \alpha_2, \dots, \alpha_r$ ($r = 2^q$)に分割して、各々の小グループとグループG_qのr個の各演算処理装置とを1対1に対応させ、小グループ α_i のなかから選択される1台の演算処理装置に対して、小グループ α_i に対応するグループG_qの演算処理装置からデータ配列Bを送信した後、小グループ α_i の演算処理装置間でデータ配列Bを送受信する操作iを、 $1 \leq i \leq r$ なるiに関して並列に実行することにより、 2^n 台の演算処理

50

装置と 2⁰ 台の演算処理装置にデータ配列 A とデータ配列 B を共有させることを特徴とする。

【0030】また、請求項 7 記載の発明は、請求項 1 乃至 6 のいずれか記載の並列計算機システムを用いて 2 台の演算処理装置間でデータを交換する工程は、演算処理装置の識別番号の大きい方から小さい方にデータを送る第 1 の送信工程と、演算処理装置の識別番号の小さい方から大きい方にデータを送る第 2 の送信工程とからなり、この第 1 の送信工程と第 2 の送信工程のうちから選択される 1 工程を先に行った後、続いて他の 1 工程を行うことを特徴とする。

【0031】

【発明の実施の形態】本発明の実施の形態について、以下、図面を参照して説明する。図 1 は並列計算機システムの構成例を示すブロック図である。ここに示した並列計算機システムは、1 台のホストの計算機 1 と 8 台の演算処理装置 2-1, 2-2, ..., 2-8 で構成されている。ホストの計算機には記憶装置 3 と通信手段 4、演算処理装置 2-1, 2-2, ..., 2-8 の各々には、個別記憶装置 5-1, 5-2, ..., 5-8 と通信手段 6-1, 6-2, ..., 6-8 が備えられている。例えば、ホストの計算機で読み込んだ入力データ等は、通信手段 4 から通信手段 6-1, 6-2, ..., 6-8 を通じて全演算処理装置に送信される。演算処理装置 2-1, 2-2, ..., 2-8 では各々割り当てられた領域の計算を行い、必要に応じて演算処理装置間の通信によりデータの授受を行う。

【0032】図 1 に示した並列計算機システムの構成に基づき、本発明にかかる並列計算機システムの第 1 の実施の形態について説明する。図 2 は本実施の形態における並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。

【0033】演算処理装置 2-1, 2-2, ..., 2-8 の識別番号をそれぞれ 0, 1, ..., 7 とし、これらを 2 進法の 3 桁の数として表示するとそれぞれ 000, 001, 010, 011, 100, 101, 110, 111 となる。8 × n 個のデータからなる配列 A が n 個のデータからなる 8 個の小配列 a₁, a₂, ..., a₈ に分割されて、8 台の演算処理装置 2-1, 2-2, ..., 2-8 に割り当てられている。それぞれの演算処理装置で割り当てられた小配列のデータに関する演算処理を行った後、配列 A の要素を全ての演算処理装置に於いて集めることを考える。なお、図 2 において各演算処理装置にかかれた 0 または 1 はそれぞれ分割された小配列を示しており、0 は計算結果が未入力の状態を、1 は計算結果が入力済みの状態を表す。

【0034】第 1 ステップとして、2⁰ の位の数を反転 (0 ならば 1, 1 ならば 0 とする) させた数を識別番号としてもつ演算処理装置との間でデータを交換する。例えば演算処理装置 0 (000) は演算処理装置 1 (001), 演算処理装置 3 (011) は演算処理装置 2 (010) と n 個のデ

ータを交換する。各演算処理装置に 2 n 個の要素が集まる。

【0035】第 2 ステップでは、2¹ の位の数を反転させた数を識別番号としてもつ演算処理装置との間でデータを交換する。例えば演算処理装置 0 (000) は演算処理装置 2 (010), 演算処理装置 3 (011) は演算処理装置 1 (001) とデータを交換する。この時、例えば演算処理装置 0 から演算処理装置 2 への送信では、演算処理装置 0 自身による演算結果の他に第 1 ステップで演算処理装置 1 から受信したデータを含む 2 n 個のデータを送信する。これにより各演算処理装置に 4 n 個の要素が集まる。

【0036】最後に第 3 ステップとして、2² の位の数を反転させた数を識別番号としてもつ演算処理装置との間でデータを交換する。例えば演算処理装置 0 (000) は演算処理装置 4 (100), 演算処理装置 3 (011) は演算処理装置 7 (111) と 4 n 個のデータを交換する。各演算処理装置に 8 n 個の要素が集まり、操作が完了する。

【0037】以上述べた通信方法は演算を 2³ = 8 個に分割した場合でありこの時のステップ数は 3 である。同様に、演算を 2⁴ = 16 個に分割し 16 台の演算処理装置において通信を行なう場合には、上述した 8 分割の場合に比べてさらに 1 ステップが必要となり、全部で 4 ステップとなる。一般に、演算を N 個に分割し N 台の演算処理装置において通信を行う場合は、上述の方法を流用して、ステップ数 log₂ N で通信が完了する。

【0038】本実施の形態の作用効果について以下検証する。例えば配列の大きさを M (word)、演算処理装置の台数を K とし、配列全体が K 分割されて各演算処理装置に渡されているものとする。K の値としては並列計算で最も一般的な条件である 2 のべき乗の場合、つまり K = 2ⁿ と表される場合について考える。この状態から、演算処理装置間の通信によって演算処理装置全部が配列全体についてデータを把握している状況を作り出すのにかかる時間について考察する。一般にデータを送信するのに要する時間 T は

$$T = A + B \times W \quad \dots\dots\dots (1)$$

と表せる。ここで、A は通信準備に要する時間で、送信するデータ量に関わらず 1 回の通信に必ず必要となる時間である。A の値はデータ量に依らない。B × W はデータ量に比例する項であり、W がデータ量 (WORD 数)、B が 1 word 当たりの転送時間である。

【0039】データの授受のステップ数は log₂ K = n である。各ステップで演算処理装置毎に送信と受信が 1 回づつ行われる。第 m ステップで授受されるデータ量は (M/K) × 2^m [word] である。データ量 M [word] のデータを全演算処理装置において集結させるのに必要な送受信の回数は各演算処理装置当たり 2 n 回であり、送受信する総データ量は

【0040】

【数 13】

$$\sum_{m=1}^n \frac{M}{K} 2^m = \frac{M}{K} 2(2^n - 1) = 2M(1 - \frac{1}{K}) \quad [\text{word}]$$

$$T(K) = 2A \log_2 K + 2M(1 - 1/K)B \quad \dots\dots\dots (2)$$

となる。

【0041】比較のため、従来法、例えば Binary tree の方式で 1 台の代表演算処理装置に全データを集めておき、同様に Binary tree の方式で全演算処理装置にデータを送信する場合の通信時間を次に求めてみる。全デー

$$T_1(K) = A \log_2 K + M(1 - 1/K)B \quad \dots\dots\dots (3)$$

となる。

【0042】代表演算処理装置から各演算処理装置にデータを配布する際のステップ数は $\log_2 K$ で、演算処理装置あたり通信回数も最大で $\log_2 K$ 回である。ただ

$$T_2(K) = A \log_2 K + (M \log_2 K)B \quad \dots\dots\dots (4)$$

となる。したがって、全通信時間 $T_0 = T_1 + T_2$ は

$$T_0(K) = 2A \log_2 K + M(1 - 1/K + \log_2 K)B \quad \dots\dots\dots (5)$$

となる。

【0043】図 3 及び図 4 のグラフは、横軸に演算処理装置台数、縦軸に通信に要する時間をとって、演算処理装置台数増加に伴う通信時間の増加の関係を示しており、従来の Binary Tree の通信方式による (5) 式の関係と、本実施の形態により通信を効率化した (2) 式の関係を、比較して示している。このグラフ中の曲線のうち実線で示した符号 10a、10b が本実施の形態の (2) 式の場合、破線で示した符号 11a、11b が従来の (5) 式の場合を示している。図 3 に示した符号 10a、11a を付した曲線は、通信されるデータ量が少なく、(1) 式の A (通信立ち上げ時間) が全通信時間 T のほぼ半分を占める状況を、また図 4 に示した符号 10b、11b を付した曲線は、通信されるデータ量が多く、(1) 式の A (通信立ち上げ時間) が全通信時間 T に比べて十分小さい状況を想定している。このグラフからも明らかなように、本実施の形態によれば、演算処理装置の台数が少数の場合、多数の場合何れも従来の方法より通信に要する時間を少なくすることができる。すなわち、本実施の形態により、データの授受の際の待ち時間を最小限に抑え、計算の高速化を図ることができる。

【0044】なお、例えば演算処理装置の台数が 16 台からなる並列計算機システムにおいて、その内の 8 台の演算処理装置の間で上述の 3 ステップからなる配列の分割分配、演算集結を行うなど、複数の演算処理装置のうち 2 の冪乗の台数だけ抜き出してこれらに通信制御用の識別番号を付与し、この台数に適応して上述した方法で配列の分割分配、演算処理を行なうものとしてもよい。

【0045】上記第 1 の実施の形態においては、関係する演算処理装置の台数が 2 の冪乗であることを前提としている。一般的な条件として演算処理装置の台数が 2 の冪乗でない場合、すなわち台数が $2^n + k$ 等として表

である。よって、本発明を適用した場合の全通信時間 T は、

タを 1 台の演算処理装置に集めるのに要する送受信の回数は、代表演算処理装置において $n = \log_2 K$ 回である。また、第 m ステップ ($m \leq n$) で送信されるデータ量は $(M/K) \times 2^{m-1}$ [word] である。よって、代表演算処理装置に全データを集めるのにかかる時間 T_1 は

し、各ステップ毎に M[word] のデータが送信される。よって、各演算処理装置にデータを配布する際にかかる時間 T_2 は

される場合にも拡張したのが以下詳述する第 2 の実施の形態である。

【0046】本発明にかかる並列計算機システムの第 2 の実施の形態について説明する。ここでは、例えば並列計算の配列を 6 分割して、6 台の演算処理装置 (識別番号を 0, 1, ..., 5 とする。) に割り当てる場合について説明する。図 5 は本実施の形態における並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。この際のデータ処理には、前記 6 台の演算処理装置のほかに 2 台の演算処理装置 (識別番号を 6, 7 とする。) を用いることとする。

【0047】第 1 ステップとして、 2^0 の位の数を反転させた数を識別番号としてもつ演算処理装置との間でデータを交換する。例えば演算処理装置 0 (000) は演算処理装置 1 (001) と、演算処理装置 3 (011) は演算処理装置 2 (010) と、それぞれ n 個のデータを交換する。演算処理装置 6 (110) と演算処理装置 7 (111) は交換すべきデータがないので休止する。この時点で、演算処理装置 0 ~ 5 に 2 n 個のデータが集められる。

【0048】第 2 ステップでは、 2^1 の位の数を反転させた数を識別番号としてもつ演算処理装置との間でデータを交換する。例えば演算処理装置 4 (100) は演算処理装置 6 (110) とのデータ交換となるが、この時点で演算処理装置 6 (110) は送信すべきデータがないので、演算処理装置 4 からデータを受信するのみとする。このデータ交換により、演算処理装置 0 ~ 3 に 4 n 個のデータが、演算処理装置 4 ~ 7 には 2 n 個のデータが集められる。

【0049】第 3 ステップでは、 2^2 の位の数を反転させた数を識別番号としてもつ演算処理装置との間でデータを交換する。例えば演算処理装置 6 (110) は演算処理装置 2 (010) との交換である。演算処理装置 6 から演算

処理装置 2 へは $2n$ 個のデータ、演算処理装置 2 から演算処理装置 6 へは $4n$ 個のデータを送信する。このようにして $6n$ 個のデータが 8 台の演算処理装置全てに行き渡る。

【0050】本実施の形態においては、一般に $(2^n + k)$ 台の演算処理装置に対して、 $(2^n - k)$ 台の演算処理装置を加えた 2^{n+1} 台の演算処理装置群を構成し、この演算処理装置群に対して上述の第 1 の実施の形態で詳述したステップにより並列計算を行うものとする。これにより、2 の冪乗ではない台数の演算処理装置に対しても 2 の冪乗の場合に準じた構成とすることで、上記第 1 の実施形態と同様の作用効果を得ることができる。

【0051】次に本発明にかかる並列計算機システムの第 3 の実施の形態を説明する。本実施の形態における演算処理装置間の通信方法について、例として、配列を 6 個の小配列分割して 6 台の演算処理装置（識別番号 0, 1, ..., 5）に割り当てている場合について説明する。まず前記配列を小配列 2 個分拡張し、拡張した部分には 0 を埋める。例えば 12 個の要素からなる配列 (3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8) であれば、4 個の要素からなる配列 (0, 0, 0, 0) を追加して、16 個の要素からなる配列 (3, 1, 4, 1, 5, 9, 2, 6, 5, 3, 5, 8, 0, 0, 0, 0) とする。演算処理装置としては前記 6 台の演算処理装置のほかに 2 台の演算処理装置（識別番号 6, 7 とする）を加えた 8 台の演算処理装置を用いる。この後は、上記第 1 の実施の形態において詳述した手順により、8 台の演算処理装置間で通信を行いデータを交換する。

【0052】本実施の形態においては、一般に $(2^n + k)$ 台の演算処理装置に対して、 $(2^n - k)$ 台の演算処理装置を加えた 2^{n+1} 台の演算処理装置群を構成し、また配列についてもその要素を 2^{n+1} 個に拡張して各演算処理装置に分配し、上記第 1 の実施の形態と同様の方法で並列計算及びデータの集結を行うものとする。これにより、2 の冪乗ではない台数の演算処理装置に対しても 2 の冪乗の場合に準じた構成とすることで、上記第 1 の実施形態と同様の作用効果を得ることができる。

【0053】次に、本発明にかかる並列計算機システムの第 4 の実施の形態について説明する。第 2 及び第 3 の実施の形態は、配列の分割数が 2 の冪でない場合、すなわち $(2^n + k)$ 個に分割される場合について、 2^{n+1} 台の演算処理装置によってデータ配列を 1 個に集結する方法について述べたものである。これに対し本実施の形態は、配列の分割数が、 $2^n + 2^m$ ($n > m$) である場合に対し、 $(2n + 2m)$ 台の演算処理装置で処理するものである。

【0054】本実施の形態における並列計算機システムの演算処理装置間の通信方法として、ここではまず例として、配列を 6 分割して 6 台の演算処理装置（識別番号 0, 1, ..., 5）に割り当てている場合について説明する。図 6 はこの場合の演算処理装置間通信方法を時系列

で示すチャートである。

【0055】まず、6 台の演算処理装置を 2 つのグループに分割する。演算処理装置グループ 1 は識別番号 0 ~ 3 の 4 台で構成される。演算処理装置グループ 2 は識別番号 4 ~ 5 の 2 台で構成される。次に、演算処理装置グループ 1 の 4 台間、および演算処理装置グループ 2 の 2 台間で、上述の第 2 の実施の形態における手順により、各々のグループでデータを集結させる。図 6 における第 1 及び第 2 ステップがこれに相当する。

【0056】この後、グループ 1 とグループ 2 でデータ交換を次の手順で行う。

[1] 演算処理装置 4 → 演算処理装置 0, 演算処理装置 5 → 演算処理装置 2 を同時並列に実施。（図 6 の第 3 ステップに相当。）

[2] 演算処理装置 1 → 演算処理装置 4, 演算処理装置 3 → 演算処理装置 5 を同時並列に実施。（図 6 の第 4 ステップに相当。）

[3] 演算処理装置 0 → 演算処理装置 1, 演算処理装置 2 → 演算処理装置 3 を同時並列に実施。（図 6 の第 5 ステップに相当。）

この方法により、6 台の演算処理装置によってデータ配列を集結させることができる。

【0057】また、本実施の形態のもう一つの例として、配列を 10 分割して 10 台の演算処理装置（識別番号 0, 1, ..., 9）に割り当てている場合について説明する。図 7 はこの場合における演算処理装置間通信方法を時系列で示すチャートである。

【0058】まず、10 台の演算処理装置を 2 つのグループに分割する。演算処理装置グループ 1 は識別番号 0, 1, ..., 7 の 8 台で構成される。演算処理装置グループ 2 は識別番号 8, 9 の 2 台で構成される。次に演算処理装置グループ 1 の 8 台の演算処理装置間、および演算処理装置グループ 2 の 2 台の演算処理装置間で、上記第 2 の実施の形態において述べた方法により、各々のグループでデータを集結させる。図 7 における第 1, 第 2 及び第 3 ステップがこれに相当する。

【0059】この後、グループ 1 とグループ 2 でデータ交換を次の手順で行う。

[1'] 演算処理装置 8 → 演算処理装置 0, 演算処理装置 9 → 演算処理装置 4 を同時並列に実施。（図 7 の第 4 ステップに相当。）

[2'] 演算処理装置 3 → 演算処理装置 8, 演算処理装置 7 → 演算処理装置 9, 演算処理装置 0 → 演算処理装置 2, 演算処理装置 4 → 演算処理装置 6 を同時並列に実施。（図 7 の第 5 ステップに相当。）

[3'] 演算処理装置 0 → 演算処理装置 1, 演算処理装置 4 → 演算処理装置 5, 演算処理装置 2 → 演算処理装置 3, 演算処理装置 6 → 演算処理装置 7 を同時並列に実施。（図 7 の第 6 ステップに相当。）

【0060】この方法により、6 台の演算処理装置によ

ってデータ配列を集結させることができる。なお、グループ2からグループ1に送信されたデータのグループ2内の分配は Binary Treeの方式によっている。

【0061】以下、本発明にかかる並列計算機システムの第5の実施の形態について説明する。本実施の形態における演算処理装置間の通信方法は、上記第5の実施の形態の通信方法を一般化したものである。以下、例として配列を22分割して22台の演算処理装置（識別番号0, 1, ..., 21）に割り当てている場合について説明する。図8及び図9はこの配列22分割の場合における演算処理装置間通信方法を時系列で示すチャートである。図8において第1ステップから第4ステップまでを、図9において第5ステップから第8ステップまでを示した。

【0062】 $22 = 2^1 + 2^2 + 2^1$ であるから、まず、演算処理装置を次の3グループに分ける。

グループ1； 識別番号0, 1, ..., 15の演算処理装置（16台）

グループ2； 識別番号16, 17, 18, 19の演算処理装置（4台）

グループ3； 識別番号20, 21の演算処理装置（2台）

【0063】次に、演算処理装置グループ1の16台間、演算処理装置グループ2の4台間、および演算処理装置グループ3の2台間で、上記第1の実施の形態の方法により各々のグループでデータを集結させる。これは図8に示した第1ステップから第4ステップまでが相当する。

【0064】この後は、上記第2或いは第3の実施の形態において説明した方法と同様の手順により、データの

グループ1の小グループ1； 演算処理装置0, 1, 2, 3
小グループ2； 演算処理装置4, 5, 6, 7
小グループ3； 演算処理装置8, 9, 10, 11
小グループ4； 演算処理装置12, 13, 14, 15

とする。

【0068】この各小グループから1台ずつ演算処理装置を選択する。ここでは演算処理装置0, 4, 8, 12を選択する。この4台の演算処理装置に対して、それぞれグループ2の演算処理装置16, 17, 18, 19から、グループ2及びグループ3に関して集結されたデータを送信する。これは図9に示した第6ステップに相当する。

【0069】次に、グループ1の各小グループにおいて、従来のBinary Treeの方式で演算処理装置間でグループ2, 3に関するデータの送受信を行ない、小グループの全演算総理装置においてグループ1, 2, 3のデータを集結させる。例えば小グループ1においては演算処理装置0から演算処理装置2に対してデータを送信し、次に演算処理装置0, 2からそれぞれ演算処理装置1, 3に対してデータの送信を行う。他の小グループにおいても同様である。これは図9に示した第7ステップ及び第8ステップに相当する。こうして、全ての22台の演算処理装置において22個のデータ配列の集結を完了する。

グループ間交換を行う。以下そのデータの通信方法を順を追って説明する。まず、第2ステップでグループ2においてデータの集結が終了するが、その時点で既にグループ3のデータの集結は完了しているから、次のステップとして、グループ2の演算処理装置16, 18とグループ3の演算処理装置19, 20との間でそれぞれデータの交換が行なわれる。これは図8に示したグループ2とグループ3における第3ステップに相当する。この時点で、グループ3の全ての演算処理装置にはグループ2及びグループ3におけるデータがすべて格納された状態となる。

【0065】次に、グループ2及びグループ3の全てのデータが格納されたグループ2の演算処理装置16, 18から、それぞれグループ2の演算処理装置17, 19に対してグループ3より受信したデータが送信される。これは図8に示したグループ2における第4ステップに相当する。

【0066】グループ1においては第4ステップで各演算処理装置間でデータの集結が終了するが、次のステップとして、グループ1と、グループ2, 3との間でデータの送受信を行う。まず、グループ1の演算処理装置0, 1, 2, 3, 4, 5から、それぞれグループ2, 3の演算処理装置16, 17, 18, 19, 20, 21に対してデータが送信される。これによりグループ2, 3においてはグループ1, 2, 3の22台の全ての演算処理装置のデータの集結が完了する。これは図9に示した第5ステップに相当する。

【0067】次に、グループ1の16台の演算処理装置を4つの小グループに分割する。すなわち、

【0070】一般に、2の冪乗では表されない台数の演算処理装置におけるデータ配列は、以上説明した方法によって集結させることができる。まず、k個の整数 $n_1, n_2, n_3, \dots, n_k$ (但し、 $n_1 > n_2 > n_3 > \dots > n_k \geq 0$)

を用いて、並列計算機システムの演算処理装置の台数を

【0071】

$$2^{n_1} + 2^{n_2} + 2^{n_3} + \dots + 2^{n_k}$$

と表す。また、データ配列をこの台数と同数の小配列に分割し、各演算処理装置に分割して演算処理を行なうものとする。並列計算機システムの演算処理装置のうち、

【0072】

【数15】

$$2^{n_1}, 2^{n_2}, 2^{n_3}, \dots, 2^{n_k}$$

台をそれぞれグループ G_1, G_2, \dots, G_k として、並列計算機システムの演算処理装置をk個のグループに分

割する。同様にデータ配列の小配列の

【0073】

【数16】

$$2^{n_1}, 2^{n_2}, 2^{n_3}, \dots, 2^{n_k}$$

個をそれぞれ配列 A_1, A_2, \dots, A_k として k 個の配列に分割する。

【0074】次に、 $1 \leq p \leq k$ なるすべての p に対して、以下の『』内に定義する操作（以下、グループ内工程 p という。）を行う。但し、グループ内工程 $1, \dots, k$ は並列して行うこととする。

【0075】『グループ G_p の（ 2 の n_p 乗）個の演算処理装置に識別番号 $0, 1, \dots, (2 \text{ の } n_p \text{ 乗} - 1)$ を付与する。次に、 $0 \leq q \leq p - 1$ なる q に対し、以下の《 》内に定義する操作 q を、 $q = 0$ から $q = p - 1$ まで順次行なう。

《 識別番号 N の演算処理装置に対し、 2 進法で表した識別番号 N の 2^q の位を反転させた番号 N' を識別番号とする演算処理装置を対応させ、データ配列の演算処理結果を識別番号 N の演算処理装置と識別番号 N' の演算処理装置との間で相互に送受信する。但し、 $q > 0$ なる q に対しては、操作 q の際に、識別番号 N, N' の演算処理装置間で、各演算処理装置による演算処理結果に加えて操作（ $q - 1$ ）までで得られた演算処理結果を合わせて送受信することとする。 》

この操作により、グループ G_p の（ 2 の n_p 乗）台の演算処理装置で、データ配列の集結を行う。』

グループの設定方法により、グループ内工程 $1, \dots, k$ を並列に行なったとき、グループ内工程 k が最初に終了し、以下、グループ内工程（ $k - 1$ ）、 $\dots, 2, 1$ の順に終了する。このことを考慮して、以下の { { } } に定義する操作（以下、グループ間工程 p という。）を、 $p = k - 1$ から $p = 1$ まで p に関して降順に行うこととする。

【0076】{ { } グループ内工程 p が終了した後、グループ G_p の各演算処理装置に集結された配列 A_p のデータを、グループ G_p の演算処理装置から、グループ G_{p+1}, \dots, G_k に属する全ての演算処理装置に送信する。すなわち、グループ G_p に属する（ 2 の n_p 乗）台の演算処理装置のうち

【0077】

【数17】

$$2^{n_{p+1}} + \dots + 2^{n_k}$$

台を選択して、これら選択された演算処理装置とグループ G_{p+1}, \dots, G_k に属する演算処理装置とを 1 対 1 に対応させ、グループ G_p からグループ G_{p+1}, \dots, G_k への配列 A_p のデータ送信を行う。次に、グループ G_{p+1} からグループ G_p へのデータの送信を行う。（ 2 の n_p 乗）台の演算処理装置からなるグループ G_p を、そ

れぞれが

【0078】

【数18】

$$2^{n_p - n_{p+1}}$$

台の演算処理装置からなる小グループ $\alpha_1, \dots, \alpha_r$ に分割する。この小グループの数 r は、

【0079】

【数19】

$$r = 2^{n_{p+1}}$$

である。ここで、グループ G_{p+1} に属する演算処理装置を b_1, \dots, b_r と表記する。グループ G_p の小グループ $\alpha_1, \dots, \alpha_r$ と、グループ G_{p+1} に属する演算処理装置を b_1, \dots, b_r とを 1 対 1 に対応させて、グループ G_{p+1} の演算処理装置 b_i から対応する小グループ α_i のうちから選択された 1 台の演算処理装置 a_i に、グループ G_{p+1} において集結された配列 A_{p+1} のデータを送信する操作を、 $1 \leq i \leq r$ なる全ての i について並列に行う。このとき、 $p < k - 1$ の場合、演算処理装置 b_i から a_i へは、グループ G_{p+2}, \dots, G_k より受信したデータ配列 A_{p+2}, \dots, A_k を含めて送信するものとする。

【0080】この後、各小グループ α_i において、演算処理装置 a_i から a_i 以外の全ての演算処理装置に対して、従来の Binary Tree の方式でデータの送信を行なう。これにより、グループ G_p の全ての演算処理装置に対してデータ配列 A_p, \dots, A_k に関するデータ配列の集結が完了する。 } }

30 この方法により、一般に複数台の演算処理装置によって各演算処理装置において分散され並列計算されたデータ配列を、効率よく集結させることができるから、計算の高速化を図ることができる。

【0081】

【発明の効果】以上説明したように本発明によれば、並列計算機システムの演算処理装置間の通信方法の効率をより向上させることにより、データの授受の際の待ち時間を最小限に抑えることができるから、並列計算機システムにおいて実施される大規模な計算の高速化を図ることができる。

【図面の簡単な説明】

【図1】本発明の第1の実施の形態における並列計算機システムの構成を示すブロック図である。

【図2】本発明の第1の実施の形態にかかる並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。

【図3】通信されるデータ量が少ない場合の本発明の第1の実施形態及び従来の通信方法を用いた場合の演算処理台数と通信時間の相関を示すグラフである。

50 【図4】通信されるデータ量が多い場合の本発明の第1

の実施形態及び従来の通信方法を用いた場合の演算処理台数と通信時間の相関を示すグラフである。

【図 5】本発明の第 2 の実施の形態にかかる並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。

【図 6】本発明の第 2 の実施の形態にかかる並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。

【図 7】本発明の第 4 の実施の形態にかかる並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。

【図 8】本発明の第 5 の実施の形態にかかる並列計算機

システムの演算処理装置間の通信方法を時系列で示すチャートである。

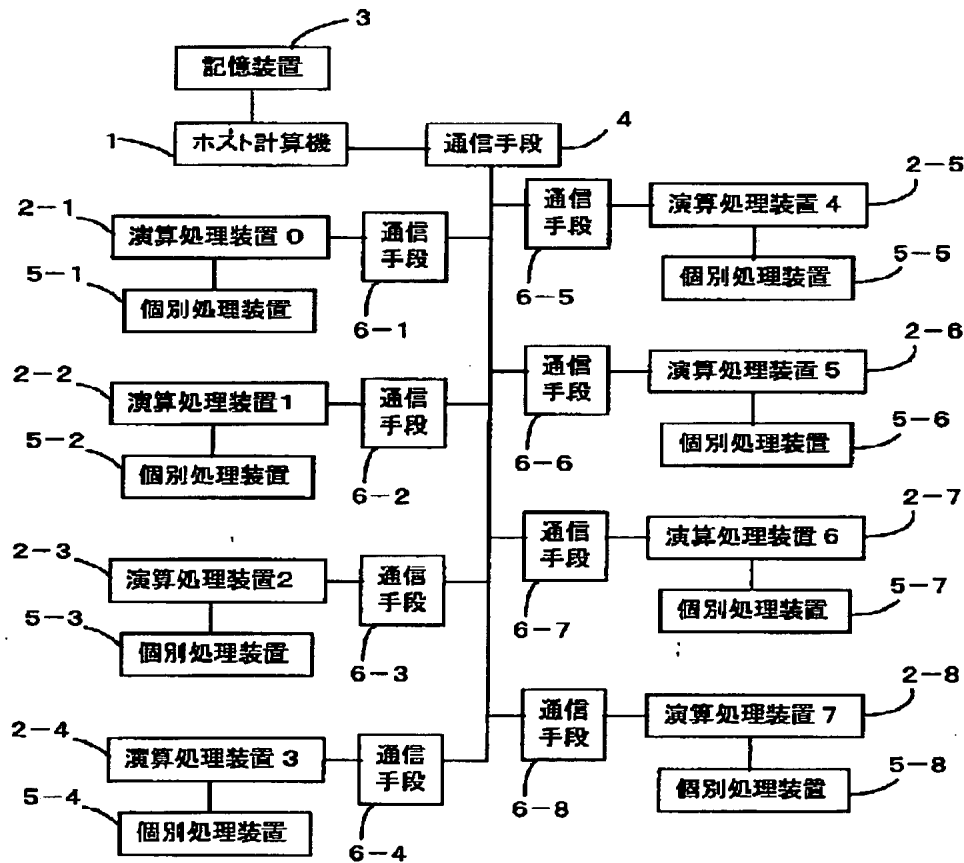
【図 9】本発明の第 5 の実施の形態にかかる並列計算機システムの演算処理装置間の通信方法を時系列で示すチャートである。

【符号の説明】

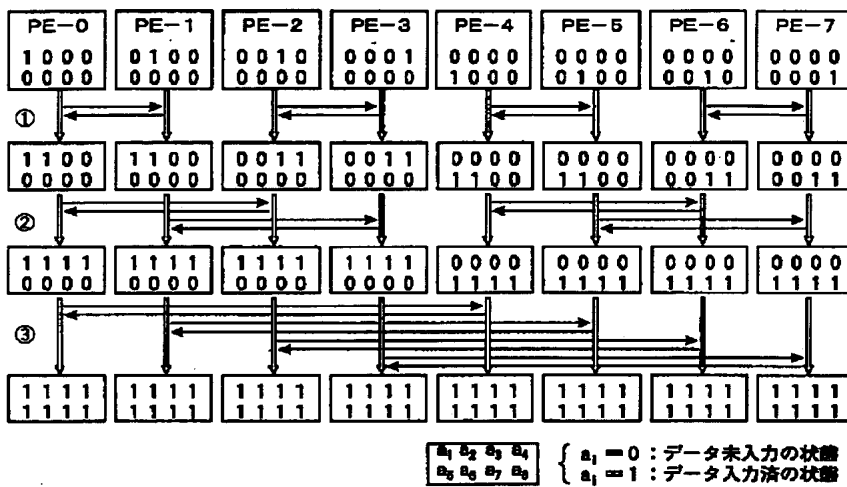
1…ホスト計算機, 2-1…演算処理装置, 3…記憶装置, 4…通信手段, 5-1…個別処理装置, 6-1…通信手段

10 a, 10 b…本発明の第 1 の実施の形態における演算処理装置の台数と通信に要する時間の関係を示す曲線

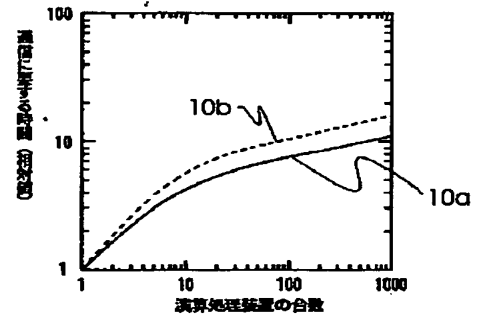
【図 1】



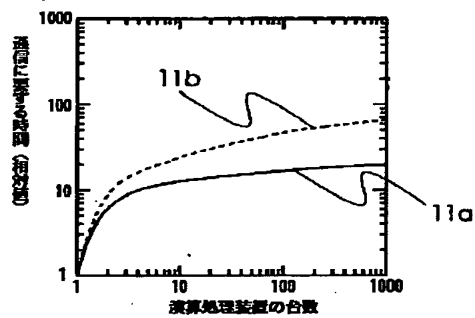
【図 2】



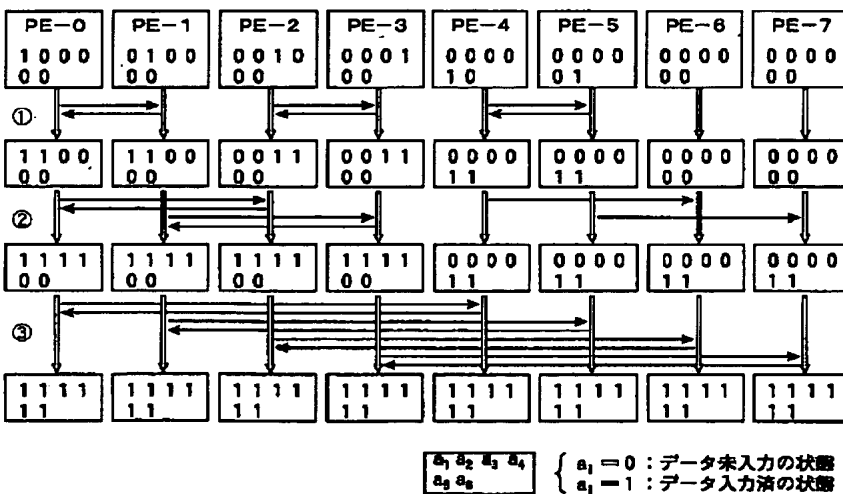
【図 3】



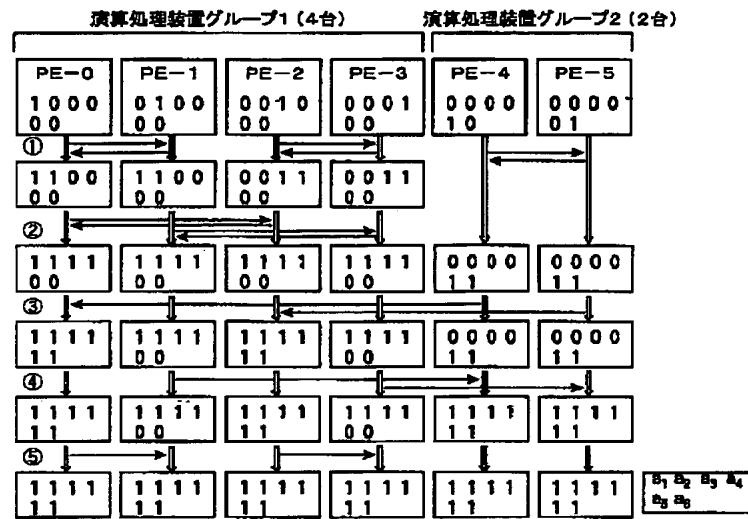
【図 4】



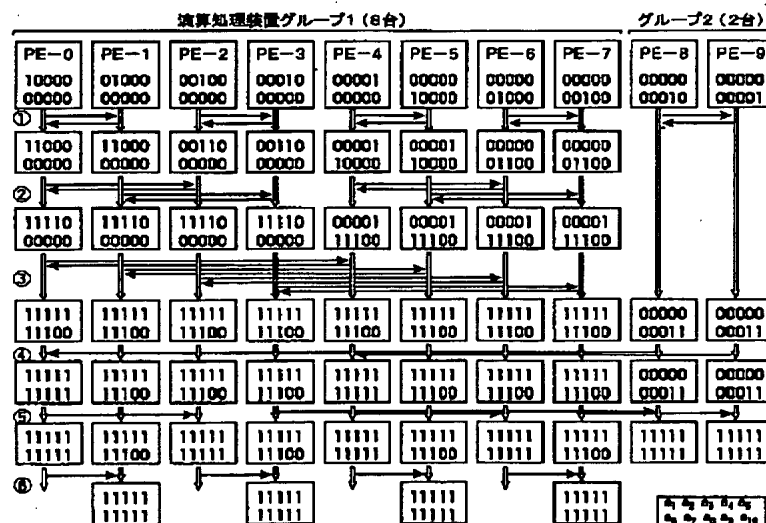
【図 5】



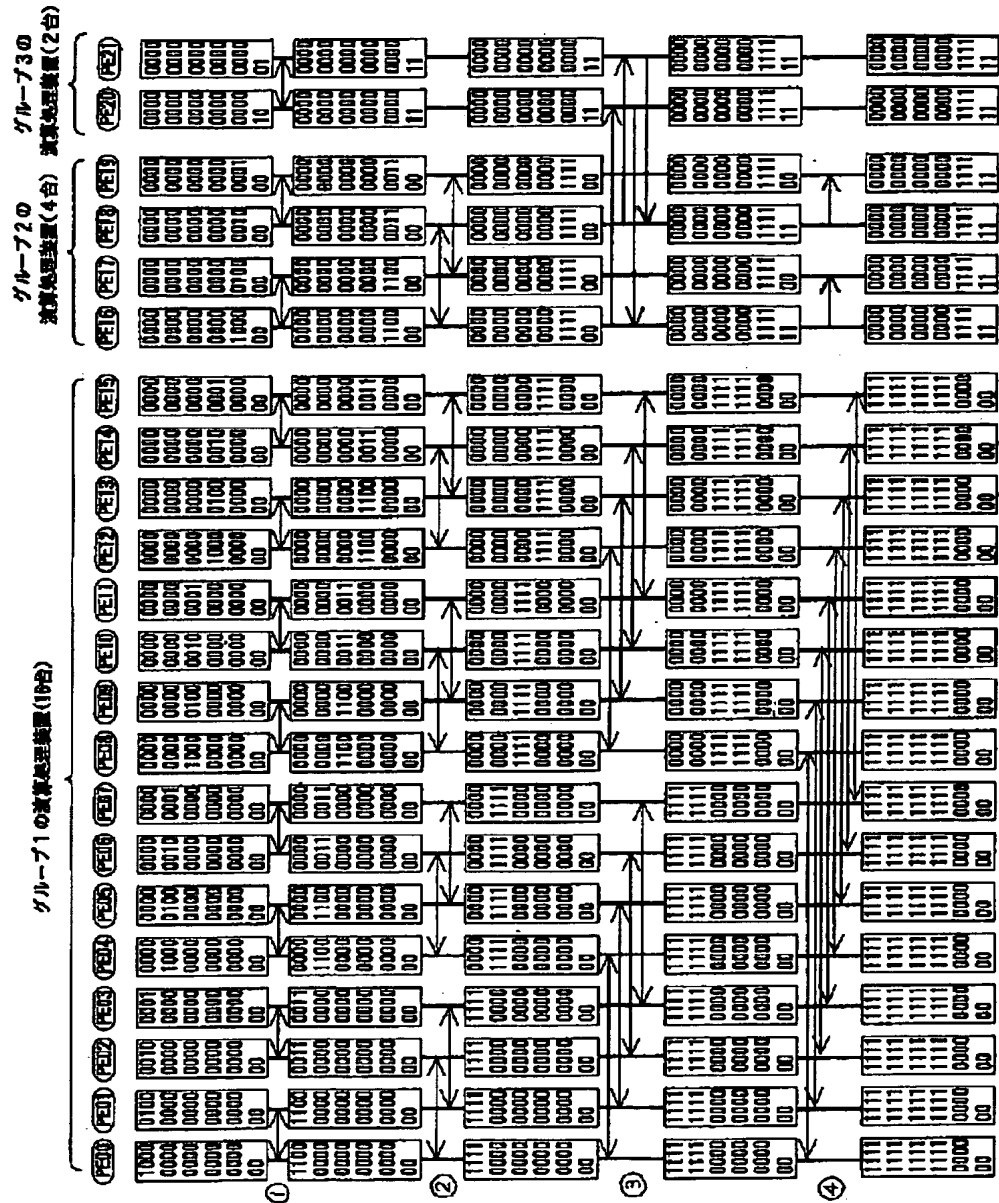
【図 6】



【図 7】



【図8】



【図 9】

